Strong converse exponents for the feedback-assisted classical capacity of entanglement-breaking channels

Dawei Ding* Mark M. Wilde†

August 4, 2016

Abstract

Quantum entanglement can be used in a communication scheme to establish a correlation between successive channel inputs that is impossible by classical means. It is known that the classical capacity of quantum channels can be enhanced by such entangled encoding schemes, but this is not always the case. In this paper, we prove that a strong converse theorem holds for the classical capacity of an entanglement-breaking channel even when it is assisted by a classical feedback link from the receiver to the transmitter. In doing so, we identify a bound on the strong converse exponent, which determines the exponentially decaying rate at which the success probability tends to zero, for a sequence of codes with communication rate exceeding capacity. Proving a strong converse, along with an achievability theorem, shows that the classical capacity is a sharp boundary between reliable and unreliable communication regimes. One of the main tools in our proof is the sandwiched Rényi relative entropy. The same method of proof is used to derive an exponential bound on the success probability when communicating over an arbitrary quantum channel assisted by classical feedback, provided that the transmitter does not use entangled encoding schemes.

1 Introduction

The classical theory of communication is one of the modern successes of applied mathematics [CT91, GK12]. It is arguably one of the foundations of our current information age and provides new ways of thinking about problems in many other fields of study, such as physics and in particular quantum mechanics. The interaction between these two fields is mutual; while some problems in quantum mechanics can be turned into communication problems, the existence of quantum phenomena strongly suggests that we should rethink many aspects of communication theory. Not only does the notion of a quantum state challenge what we mean by "information," but the possibilities due to quantum mechanics give rise to new classes of communication protocols. With respect to this latter consideration, some fundamental motivating questions for quantum information theory have traditionally been and still are the following:

1. Is the theory of classical communication affected at a fundamental level by the consideration of quantum mechanical phenomena?

*Department of Applied Physics, Stanford University, Stanford, California 94305-4090, USA

†Hearne Institute for Theoretical Physics, Department of Physics and Astronomy, Center for Computation and Technology, Louisiana State University, Baton Rouge, Louisiana 70803, USA

2. Does using quantum states and measurement for classical communication have practical advantages over using classical techniques?

In an attempt to answer these questions, one of the primary goals is to study the ability of a quantum channel to communicate classical information, that is, bits. Like many communication problems, this ability is quantified by the notion of channel capacity. The *classical capacity* C of a quantum channel $\mathcal{N}$ is defined to be the maximum rate of communication such that the decoding error probability can tend to zero in the limit of many channel uses. With this definition, one natural question is to determine how to compute the classical capacity. We know that the Holevo-Schumacher-Westmoreland (HSW) theorem [Hol98, SW97] provides a lower bound:

$$C(\mathcal{N}) \geqslant \chi(\mathcal{N}) \equiv \sup_{\{p_X(x),\rho_x\}} I(X;B)_\rho, \tag{1.1}$$

where $\{p_X(x), \rho_x\}$ is an ensemble of quantum states such that each ρ_x can be input to the channel, and $I(X;B)_\rho \equiv H(X)_\rho + H(B)_\rho - H(XB)_\rho$ is the quantum mutual information of the following classical-quantum state:

$$\rho_{XB} \equiv \sum_x p_X(x)|x\rangle\langle x|_X \otimes \mathcal{N}(\rho_x), \tag{1.2}$$

where $\{|x\rangle\}$ is an orthonormal basis for the classical reference system and $H(G)_\sigma$ is the von Neumann entropy of a quantum state σ_G on system G. The quantity $\chi(\mathcal{N})$ is called the *Holevo information* of the channel.

The classical capacity $C(\mathcal{N})$ can actually be formally rewritten in terms of the Holevo information as well, via a procedure known as regularization. Before doing so, note that we obtain the HSW theorem by considering only encoding procedures that do not use entangled inputs[1]. That is, each quantum system sent through the channel is not entangled with any other system that is sent. To get the classical capacity, it is generally necessary to incorporate entangled inputs into the calculation. The approach given by [Hol98, SW97] is to multiplex the channel such that one use of this multiplexed channel corresponds to multiple uses of the original channel. A multipartite entangled state describing several inputs across different uses of the original channel can now be simulated by the corresponding single input state to the multiplexed channel. We can therefore express the maximum rate for blocks of size n in terms of a Holevo information:

$$\frac{1}{n}\chi(\mathcal{N}^{\otimes n}). \tag{1.3}$$

To obtain the classical capacity, we simply allow for inputs entangled across arbitrarily many channel uses:

$$C(\mathcal{N}) = \lim_{n\to\infty} \frac{1}{n}\chi(\mathcal{N}^{\otimes n}). \tag{1.4}$$

This idea put together with a converse theorem establishes the regularized expression in (1.4) as being equal to the classical capacity.

Unfortunately, computing the classical capacity this way is clearly intractable. This prompts us to look for special cases. We observe that the limit in (1.4) is equal to $\chi(\mathcal{N})$ iff the Holevo information satisfies a tensor-power additivity property (see Appendix A for a brief derivation):

$$\forall n \quad \chi(\mathcal{N}^{\otimes n}) = n\chi(\mathcal{N}). \tag{1.5}$$

[1]The highest possible rate with this restriction, proven by the HSW theorem to be $\chi(\mathcal{N})$, is then a lower bound on the classical capacity.

Another way to restate the above is that entangled inputs do not increase capacity iff (1.5) is satisfied. However, this is not true for some channels, as was shown in [Has09]. The next question then appears: What characterizes channels that satisfy (1.5)?

A sufficient condition for a quantum channel to satisfy tensor-power additivity is for it to be entanglement-breaking (EB) [HSR03]. Let $\mathcal{N}_{A\to B}$ denote a quantum channel, where the arrow notation indicates that the channel maps a state of the input system A to a state in an output system B. An EB channel is defined such that for any bipartite state $\rho_{AA'}$, the output state

$$\left(\mathcal{N}_{A\to B}\otimes \mathrm{id}_{A'}\right)\left(\rho_{AA'}\right) \tag{1.6}$$

is separable. That is, the output of an EB channel can be written as a convex sum of product states. Effectively, the channel "breaks" the entanglement between A and A'. Previous results have established that the Holevo capacity of EB channels is additive[2] [Sho02]. This is intuitive since any entanglement of the inputs is broken by the channel. Following this line of thinking, one could consider making a stronger statement by allowing additional resources to assist the communication but not to the point that entanglement can be established. Indeed, [BN05] proves a generalization of tensor-power additivity for EB channels with noiseless classical feedback. Furthermore, their results show that classical feedback does not increase the capacity of EB channels.

There are also possible stronger statements in another direction. The direct part of the original HSW theorem states that if the rate R of communication is less than the Holevo information χ, then there exists a sequence of protocols $\mathcal{P}^n$ such that the probability of error for such a sequence satisfies

$$\lim_{n\to\infty} p_e(n) = 0, \tag{1.7}$$

where n is the number of channel uses. A result from [Hay07] sharpens this claim by showing that there exists a sequence $\mathcal{P}^n$ such that

$$p_e(n) \leqslant 2^{-kn}, \tag{1.8}$$

for some $k > 0$ determined by the channel. The converse part of the original HSW theorem can be strengthened in a similar manner. It states that if $R > C(\mathcal{N})$, for any sequence of protocols $\mathcal{P}^n$, then

$$\lim_{n\to\infty} p_e(n) > 0. \tag{1.9}$$

This is known as the *weak converse*. In contrast, a *strong converse* is symmetric to the achievability result above and states that regardless of the protocols used, the success probability decreases to zero in the asymptotic limit whenever $R > C(\mathcal{N})$. The strong converse can be sharpened as well whenever there is a constant separation between R and $C(\mathcal{N})$, such that the convergence of the success probability to zero is exponential in n.

There are many reasons why we would want to prove a strong converse. First, a strong converse enriches our understanding of the capacity. A strong converse along with an achievability theorem shows that the capacity is a sharp boundary between reliable and unreliable communication regimes. This, amongst other results, indicates that the classical capacity of a quantum channel is a fundamental quantity of interest. Second, a strong converse is more relevant in practice than is the weak converse. A realistic quantum communication scheme has a finite blocklength; that is,

[2]Shor proved that the Holevo capacity of a tensor product of an EB channel with any other channel is equal to the sum of their respective Holevo informations. This form of additivity is stronger than (1.5) and is the one usually found in the literature.

the encoding is across a finite number of channel uses. Although the weak converse does provide a lower bound on the probability of error in the non-asymptotic regime, the strong converse improves the bound considerably. It expresses a trade-off between rate, error probability, and blocklength restrictive enough to be easily checked numerically or experimentally.

While a classical version of the strong converse is known for arbitrary discrete memoryless classical channels [Wol78, Ari73], it is still open whether or not strong converses hold for memoryless quantum channels. After some early work [ON99, Win99], it has been proved for special cases, in particular for channels with certain symmetry [KW09], for EB channels [WWY14], and for a wide class of quantum Gaussian channels [BGPWW15]. Given the strong converse results for EB channels [WWY14] and the weak converse for EB channels with feedback [BN05], it is natural to ask if these two statements are true at the same time. We can also ask directly for the strong converse for unentangled inputs when a feedback link is available. These are the main questions that we address in this paper.

2 Summary of Results

In this paper, we derive an explicit exponential bound on the success probability of a classical communication scheme that uses an entanglement-breaking channel along with classical feedback. The same method of proof can be used to establish a bound for arbitrary quantum channels with classical feedback, provided that the inputs are not entangled across multiple uses of the channel. When the communication rate exceeds the classical capacity, these exponential bounds immediately imply strong converse theorems for these settings.

We now provide an outline of the proof:

1. First, taking as a starting point the approach of [Nag01], relating hypothesis testing to unassisted communication, we bound the success probability of an arbitrary feedback-assisted classical communication protocol by a sandwiched Rényi relative entropy [MLDS$^+$13, WWY14].

2. Next, one of the main observations from [BN05] is that the sender and receiver's systems are separable at all times throughout such a protocol, whenever the communication channel is entanglement-breaking. We use this fact and an entropy inequality from [Kin03] to split the relative entropy into two terms. The first term is bounded by an α-information radius, which is a measure of the range of states a channel can output. We then equate this to the sandwiched α-Holevo information [WWY14], which is a Rényi generalization of the Holevo information. The second term is bounded via monotonicity by another sandwiched Rényi relative entropy, to which we recursively apply the same argument.

3. This gives the following bound on the probability of success for any finite blocklength n:

$$p_{\text{succ}} \leqslant 2^{-n \sup_{\alpha>1} \frac{\alpha-1}{\alpha}(R-\widetilde{\chi}_\alpha(\mathcal{N}))}, \tag{2.1}$$

   where R is the rate of communication and $\mathcal{N}$ is the EB channel. It follows from previous arguments [WWY14] that when $R > \chi(\mathcal{N})$, the right hand side of (2.1) is a decaying exponential, thereby establishing the strong converse. We provide an alternate (arguably simpler) proof of this fact (similar to those in [MH11, CMW16]) by establishing that the α-Holevo information converges continuously to the Holevo information as α approaches 1.

Appendix B includes a brief review of the argument for the weak converse from [BN05].

3 Preliminaries

In this section we provide some necessary definitions, concepts, and previous results used in the derivation of (2.1).

3.1 Quantum states, measurements, operator norms, and quantum channels

We start with definitions of relevant mathematical notions from quantum mechanics. Given a finite-dimensional Hilbert space $\mathcal{H}$, let $\mathcal{B}(\mathcal{H})$ denote the algebra consisting of bounded linear operators acting on $\mathcal{H}$. A relevant measure of a bounded operator X is its *Schatten α-norm*, which is defined as

$$\|X\|_\alpha \equiv \left\{\mathrm{Tr}\left[|X|^\alpha\right]\right\}^{1/\alpha}, \tag{3.1}$$

where $\alpha \geqslant 1$ and $|X| \equiv \sqrt{X^\dagger X}$.

Now, let $\mathcal{T}(\mathcal{H}) \subseteq \mathcal{B}(\mathcal{H})$ denote the subset of trace-class operators. The set of *quantum states* is a convex subset of $\mathcal{T}(\mathcal{H})$ given by

$$\mathcal{S}(\mathcal{H}) = \left\{\rho \in \mathcal{T}(\mathcal{H}) : \rho^\dagger = \rho, \rho \geqslant 0, \mathrm{Tr}\,\rho = 1\right\}, \tag{3.2}$$

where the notation $\rho \geqslant 0$ means that ρ is positive semidefinite. For composite states, we consider the tensor product of two Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, denoted by $\mathcal{H}_A \otimes \mathcal{H}_B$. We can obtain from the overall density operator $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ the reduced density operator ρ_A corresponding to the quantum state on only the A system by performing a partial trace:

$$\rho_A = \mathrm{Tr}_B(\rho_{AB}) \equiv \sum_b (I_A \otimes \langle b|_B)\rho_{AB}(I_A \otimes |b\rangle_B) \tag{3.3}$$

where $\{|b\rangle_B\}$ is an orthonormal basis of states on the B system. A state ρ_{AB} is *separable* if it can be written as

$$\rho_{AB} = \sum_x p(x)\rho_A^x \otimes \rho_B^x, \tag{3.4}$$

where $p(x)$ is a probability distribution and $\{\rho_A^x\}$ and $\{\rho_B^x\}$ are sets of states.

A positive operator-valued measure (POVM) consists of a set $\{\Lambda_m\}$ of positive semidefinite operators indexed by m and corresponding to different measurement results. The set satisfies $\sum_m \Lambda_m = I$, which allows us to interpret the quantity

$$p_m \equiv \mathrm{Tr}(\rho\Lambda_m) \tag{3.5}$$

as the probability of measuring m given a quantum state ρ.

We next consider maps on trace class operators and in particular quantum states. A linear map $\Psi : \mathcal{T}(\mathcal{H}_A) \to \mathcal{T}(\mathcal{H}_B)$ is called *positive* if $\forall \tau \in \mathcal{T}(\mathcal{H}_A)$, $\tau \geqslant 0$ implies $\Psi(\tau) \geqslant 0$. It is called *completely positive* if $\mathrm{id}_R \otimes \Psi$ is positive for an arbitrary auxiliary system R, where id_R is the identity map on R. A map Ψ is a *quantum channel* if it is linear, completely positive, and also trace-preserving.

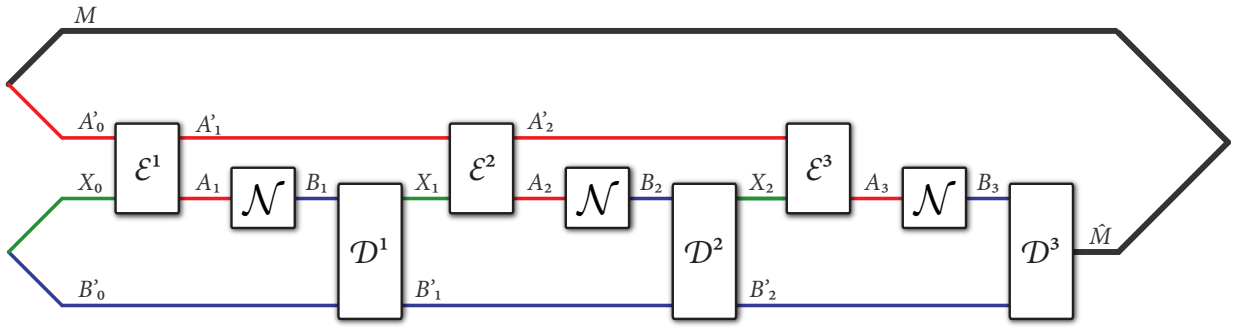

Figure 1: A general protocol for feedback-assisted classical communication. The sender manipulates the systems labeled by A and the receiver those labeled by B. Every X system is classical and represents noiseless classical feedback from the receiver to the sender.

3.2 Classical feedback-assisted classical communication protocols

Figure 1 depicts the most general three-round protocol for classical feedback-assisted classical communication. The generalization to n rounds is clear. The protocol begins with Alice preparing a classical register M with the message to be sent. This is correlated with some system A'_0. Bob uses the classical feedback channel to send a classical system X_0 correlated with some quantum system B'_0 to Alice. The global state is then

$$\rho_{MA'_0X_0B'_0} \equiv \sum_m p_M(m) \left|m\right\rangle\!\left\langle m\right|_M \otimes \rho^m_{A'_0} \otimes \sum_x p_{X_0}(x) \left|x\right\rangle\!\left\langle x\right|_{X_0} \otimes \rho^x_{B'_0}. \tag{3.6}$$

We will track the state conditioned on a particular value m of the message register M, given by

$$\rho^m_{A'_0X_0B'_0} \equiv \rho^m_{A'_0} \otimes \sum_x p_{X_0}(x) \left|x\right\rangle\!\left\langle x\right|_{X_0} \otimes \rho^x_{B'_0}. \tag{3.7}$$

Alice performs an encoding $\mathcal{E}^1_{A'_0X_0\to A'_1A_1}$. The state at this point is

$$\rho^m_{A'_1A_1B'_0} \equiv \mathcal{E}^1_{A'_0X_0\to A'_1A_1}\left(\rho^m_{A'_0X_0B'_0}\right) \tag{3.8}$$

$$= \sum_x p_{X_0}(x) \mathcal{E}^1_{A'_0X_0\to A'_1A_1}\left(\rho^m_{A'_0} \otimes \left|x\right\rangle\!\left\langle x\right|_{X_0}\right) \otimes \rho^x_{B'_0}, \tag{3.9}$$

We note that the state $\rho^m_{A'_1A_1B'_0}$ is separable with respect to the cut $A'_1A_1 : B'_0$.

Next, Alice uses the EB channel $\mathcal{N}_{A_1\to B_1}$ for the first time by sending system A_1, leading to the state

$$\rho^m_{A'_1B_1B'_0} \equiv \mathcal{N}_{A_1\to B_1}\left(\rho^m_{A'_1A_1B'_0}\right). \tag{3.10}$$

Since the channel is entanglement-breaking and the state before the channel was already separable with respect to the $A'_1A_1 : B'_0$ cut, the state after the channel is fully separable. Now, consider instead that Alice uses an arbitrary channel $\mathcal{N}_{A_1\to B_1}$ but with separable inputs. Then, the state $\rho^m_{A'_1B_1B'_0}$ is again fully separable. We write it in the form

$$\rho^m_{A'_1B_1B'_0} = \sum_y p_Y(y) \rho^{y,m}_{A'_1} \otimes \rho^{y,m}_{B_1} \otimes \rho^{y,m}_{B'_0}. \tag{3.11}$$

Bob now applies the decoding map $\mathcal{D}^1_{B_1B_0'\to X_1B_1'}$, where X_1 is the classical system that is sent back to Alice. The state at this point is

$$\rho^m_{A_1'X_1B_1'} \equiv \mathcal{D}^1_{B_1B_0'\to X_1B_1'}\left(\rho^m_{A_1'B_1B_0'}\right) \tag{3.12}$$

$$= \sum_{x_1} p_{X_1}(x_1)\,|x_1\rangle\langle x_1|_{X_1} \otimes \sum_{z_1} p_{Z_1|X_1}(z_1|x_1)\rho^{x_1,z_1}_{A_1'} \otimes \rho^{x_1,z_1}_{B_1'}. \tag{3.13}$$

which is fully separable. Hence, after Alice applies the second encoder $\mathcal{E}^2$ and then sends it through the channel, the state will still be fully separable. The key observation here is that if we use an EB channel or an arbitrary channel with separable inputs, *the state is always separable across a cut that divides Alice and Bob's systems.*

The only difference in subsequent rounds is the final measurement. Say there are n rounds in the protocol. At the last round, Bob measures the state $\rho_{B_nB'_{n-1}}$ using a POVM given by $\{D^m\}$ with elements corresponding to different possible messages that Alice sent.

3.3 Rényi relative entropies and bounds on success probability

An important classical information theoretic quantity is the Rényi relative entropy, which can be generalized to the quantum case in a number of ways. In this paper, we use the *sandwiched quantum Rényi relative entropy* [MLDS+13, WWY14] which is given by

$$\widetilde{D}_\alpha(\rho\|\sigma) \equiv \begin{cases} \frac{1}{\alpha-1}\log\left[\mathrm{Tr}\left(\left(\sigma^{(1-\alpha)/(2\alpha)}\rho\sigma^{(1-\alpha)/(2\alpha)}\right)^\alpha\right)\right] & : \rho \not\perp \sigma \wedge (\mathrm{supp}(\rho)\subseteq\mathrm{supp}(\sigma) \vee \alpha\in(0,1)) \\ +\infty & : \text{otherwise} \end{cases} \tag{3.14}$$

where $\alpha \in (0,1)\cup(1,\infty)$ and $\rho \not\perp \sigma$ means ρ,σ are non-orthogonal quantum states. Note that all logarithms in this paper are taken base two.

We now recall some properties of the sandwiched Rényi relative entropy. For fixed ρ and σ, the function $\alpha \mapsto \widetilde{D}_\alpha(\rho\|\sigma)$ is monotone non-decreasing [MLDS+13]. It also converges to the quantum relative entropy $D(\rho\|\sigma)$ [Ume62] in the limit as $\alpha\to 1$ [MLDS+13, WWY14]:

$$\lim_{\alpha\to 1}\widetilde{D}_\alpha(\rho\|\sigma) = D(\rho\|\sigma), \tag{3.15}$$

where

$$D(\rho\|\sigma) \equiv \begin{cases} \mathrm{Tr}\left[\rho\left(\log\rho - \log\sigma\right)\right] & : \mathrm{supp}(\rho)\subseteq\mathrm{supp}(\sigma) \\ +\infty & : \text{otherwise} \end{cases}. \tag{3.16}$$

Furthermore, it satisfies the data-processing inequality for $\alpha\in[1/2,1)\cup(1,\infty)$ [FL13, Bei13]; that is, for all quantum channels $\mathcal{N}$,

$$\widetilde{D}_\alpha(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \leqslant \widetilde{D}_\alpha(\rho\|\sigma). \tag{3.17}$$

In particular, consider the following replacement channel which simply replaces the input with some state ω:

$$\mathcal{R}_\omega(\rho) \equiv \mathrm{Tr}(\rho)\omega. \tag{3.18}$$

It is easy to see that for all ω, $\widetilde{D}_\alpha(\omega\|\omega) = 0$, implying that

$$\widetilde{D}_\alpha(\rho\|\sigma) \geqslant \widetilde{D}_\alpha(\mathcal{R}_\omega(\rho)\|\mathcal{R}_\omega(\sigma)) = 0, \tag{3.19}$$

which shows that the relative entropy is non-negative whenever its arguments are quantum states ρ and σ.

Using the relative entropy, we can define the sandwiched α-*Holevo information* [WWY14] of an ensemble, that is, a classical probability distribution of quantum states, $\{p_X(x), \rho_x\}$ as

$$\widetilde{\chi}_\alpha(\{p_X(x), \rho_x\}) \equiv \inf_{\sigma_R \in \mathcal{S}(\mathcal{H}_R)} \widetilde{D}_\alpha(\rho_{XR} \| \rho_X \otimes \sigma_R) \tag{3.20}$$

where

$$\rho_{XR} \equiv \sum_x p_X(x) |x\rangle\langle x|_X \otimes (\rho_x)_R \tag{3.21}$$

and ρ_x are states of a system R. With this, we define the α-Holevo information of a quantum channel $\mathcal{N}$ as

$$\widetilde{\chi}_\alpha(\mathcal{N}) \equiv \sup_{\{p_X(x), \rho_x\}} \widetilde{\chi}_\alpha\left(\{p_X(x), \mathcal{N}(\rho_x)\}\right). \tag{3.22}$$

We also define the α-information radius of a channel $\mathcal{N}$ as

$$\widetilde{K}_\alpha(\mathcal{N}) \equiv \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \sup_{\rho \in \mathcal{S}(\mathcal{H})} \widetilde{D}_\alpha(\mathcal{N}(\rho) \| \sigma). \tag{3.23}$$

We note that since $\widetilde{D}_\alpha$ is monotonically non-decreasing in α, so are $\widetilde{\chi}_\alpha$ and $\widetilde{K}_\alpha$.

4 Strong converse

This section is dedicated to proving the main theorem of this paper. We will need the following lemmas. The first is proven in [Nag01] and in Lemma 5 of [WWY14] via monotonicity of $\widetilde{D}_\alpha$. The second will be used to take advantage of the separability of the quantum state observed in Section 3.2. The third is an equality between α-Holevo information and α-information radius. The fourth shows that α-Holevo information and α-information radius tends to the conventional Holevo information and information radius in the limit $\alpha \to 1$. Note that this establishes Lemma 3 as a generalization of the equality $\chi(\mathcal{N}) = K(\mathcal{N})$ [OPW97, SW01, SW02].

Lemma 1. *Let $\alpha > 1$, $\rho, \sigma \in \mathcal{S}(\mathcal{H})$, and Λ be such that $0 \leqslant \Lambda \leqslant I$. Let*

$$p \equiv \operatorname{Tr}\left[\Lambda \rho\right], \quad q \equiv \operatorname{Tr}\left[\Lambda \sigma\right]. \tag{4.1}$$

Then

$$\widetilde{D}_\alpha(\rho \| \sigma) \geqslant \frac{1}{\alpha - 1} \log\left[p^\alpha q^{1-\alpha}\right]. \tag{4.2}$$

Lemma 2 ([Kin03, Hol06]). *Let P_{AB} be a positive semidefinite separable operator. Such an operator can be written in the following form:*

$$P_{AB} = \sum_j C_A^j \otimes D_B^j, \tag{4.3}$$

where $C_A^j, D_B^j \geqslant 0$ for all j. Let $P_B = \operatorname{Tr}_A\{P_{AB}\}$ and let $\mathcal{M}_A$ be a completely positive linear map acting on the A system. Then, for all $\alpha \geqslant 1$,

$$\left\|\left(\mathcal{M}_A \otimes \mathrm{id}_B\right)(P_{AB})\right\|_\alpha \leqslant \nu_\alpha(\mathcal{M}_A) \cdot \left\|P_B\right\|_\alpha, \tag{4.4}$$

where $\nu_\alpha(\mathcal{M}_A)$ is the $1 \to \alpha$ norm of $\mathcal{M}_A$, defined as

$$\nu_\alpha(\mathcal{M}_A) \equiv \sup_{X \neq 0, X \in \mathcal{T}(\mathcal{H}_A)} \frac{\|\mathcal{M}_A(X)\|_\alpha}{\|X\|_1}. \tag{4.5}$$

Lemma 3 ([WWY14]). *For $\alpha > 1$, the α-Holevo information and the α-information radius are the same*[3]

$$\widetilde{\chi}_\alpha(\mathcal{N}) = \widetilde{K}_\alpha(\mathcal{N}). \tag{4.6}$$

Lemma 4. *For a quantum channel $\mathcal{N}$, the following limits hold:*

$$\lim_{\alpha \to 1} \widetilde{\chi}_\alpha(\mathcal{N}) = \chi(\mathcal{N}) \tag{4.7}$$

and

$$\lim_{\alpha \to 1} \widetilde{K}_\alpha(\mathcal{N}) = K(\mathcal{N}), \tag{4.8}$$

where

$$K(\mathcal{N}) \equiv \inf_{\sigma \in \mathcal{S}(\mathcal{H})} \sup_{\rho \in \mathcal{S}(\mathcal{H})} D(\mathcal{N}(\rho) \| \sigma) \tag{4.9}$$

is the information radius[4]*.*

We now state the theorem.

Theorem 5. *Given any n-round protocol for classical feedback-assisted classical communication over an entanglement-breaking channel $\mathcal{N}$ with rate R, the average probability of success is bounded from above by an exponential in n:*

$$p_{\text{succ}} \leqslant 2^{-n \sup_{\alpha > 1} \frac{\alpha - 1}{\alpha}(R - \widetilde{\chi}_\alpha(\mathcal{N}))}, \tag{4.10}$$

where $\widetilde{\chi}_\alpha$ is the α-Holevo information. The same bound holds for an arbitrary channel $\mathcal{N}$ given that the encoder does not entangle inputs across different uses of the channel.

Proof. We take as a starting point the approach of Nagaoka [Nag01], connecting hypothesis testing with data processing of a Rényi information quantity. Let $\mathcal{P}_n$ be such a protocol. We are bounding the average probability of success, so we assume Alice chooses her messages uniformly at random. Using the notation of Section 3.2, the state we have at the final round of the protocol $\mathcal{P}_n$ is

$$\rho_{M B_n B'_{n-1}} = \frac{1}{L} \sum_{m=1}^{L} |m\rangle\langle m|_M \otimes \rho^m_{B_n B'_{n-1}}, \tag{4.11}$$

where L is the number of possible messages.

Following the argument in [CMW16], we can write the success probability as

$$p_{\text{succ}} = \frac{1}{L} \sum_m \operatorname{Tr}\left[D^m_{B_n B'_{n-1}} \rho^m_{B_n B'_{n-1}} \right] = \operatorname{Tr}\left[T_{M B_n B'_{n-1}} \rho_{M B_n B'_{n-1}} \right], \tag{4.12}$$

[3] To get all $\alpha > 1$, we actually need to extend the lemma in [WWY14] slightly. This follows from the proof given there and the observations that for all $\alpha > 1$, $x^{(1-\alpha)/\alpha}$ is operator convex, $\operatorname{Tr}\{x^\alpha\}$ is convex, and $\widetilde{D}_\alpha$ is jointly quasi-convex.

[4] This lemma is proved in Appendix C.

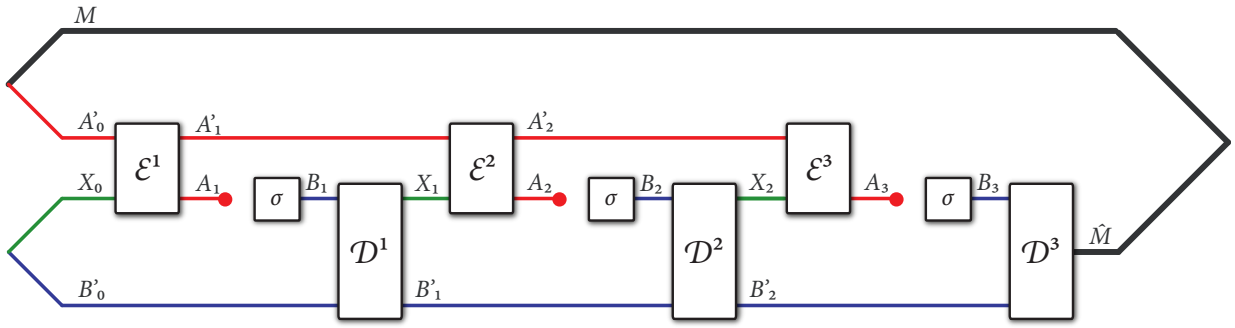

Figure 2: A protocol for feedback-assisted classical communication when using the replacement channel. Notice that the communication line between A_1 and $\hat{M}$ is broken.

where

$$T_{MB_nB'_{n-1}} \equiv \sum_m |m\rangle\langle m|_M \otimes D^m_{B_nB'_{n-1}}. \tag{4.13}$$

Note that $0 \leqslant T \leqslant I$, so that $\{T, I-T\}$ is a POVM.

We now consider the state $\tau_{MB_nB'_{n-1}}$ defined to be the final state if we had implemented $\mathcal{P}_n$ using a replacement channel $\mathcal{R}_\sigma$ instead of the original channel (see Figure 2). The overall state in this alternate scenario has a simple expression since Bob's states are now independent of m:

$$\tau_{MB_nB'_{n-1}} = \tau_M \otimes \sigma_{B_n} \otimes \tau_{B'_{n-1}}, \tag{4.14}$$

where $\tau_M = I_M/L$. This allows us to compute the following:

$$\operatorname{Tr}\left[T_{MB_nB'_{n-1}}\tau_{MB_nB'_{n-1}}\right] = \frac{1}{L}\sum_m \operatorname{Tr}\left[D^m_{B_nB'_{n-1}}\sigma_{B_n}\otimes\tau_{B'_{n-1}}\right] = \frac{1}{L}, \tag{4.15}$$

where the last equality follows because $\sum_m D^m = I$ and $\operatorname{Tr}[\sigma_{B_n}\otimes\tau_{B'_{n-1}}] = 1$. This equality is intuitive, since for a replacement channel, the receiver Bob cannot do any better than to guess the input message m at random. Letting T be the operator Λ in Lemma 1, we conclude that for $\alpha > 1$,

$$\widetilde{D}_\alpha(\rho_{MB_nB'_{n-1}}\|\tau_{MB_nB'_{n-1}}) \geqslant \frac{1}{\alpha-1}\log\left[p^\alpha_{\text{succ}}\frac{1}{L^{1-\alpha}}\right], \tag{4.16}$$

which can be re-written as follows:

$$\frac{\alpha}{\alpha-1}\log(p_{\text{succ}}) + \log L \leqslant \widetilde{D}_\alpha(\rho_{MB_nB'_{n-1}}\|\tau_{MB_nB'_{n-1}}). \tag{4.17}$$

We would therefore like to bound $\widetilde{D}_\alpha(\rho_{MB_nB'_{n-1}}\|\tau_{MB_nB'_{n-1}})$. To do so, consider that

$$\begin{aligned}\widetilde{D}_\alpha(\rho_{MB_nB'_{n-1}}\|\tau_{MB_nB'_{n-1}}) &= \widetilde{D}_\alpha(\mathcal{N}_{A_n\to B_n}(\rho_{MA_nB'_{n-1}})\|\tau_{MB'_{n-1}}\otimes\sigma_{B_n}) \\ &= \frac{\alpha}{\alpha-1}\log\left\|\left(\Theta_{\sigma_{B_n}^{\frac{1-\alpha}{\alpha}}}\circ\mathcal{N}_{A_n\to B_n}\right)\left(\tau^{(1-\alpha)/(2\alpha)}_{MB'_{n-1}}\rho_{MA_nB'_{n-1}}\tau^{(1-\alpha)/(2\alpha)}_{MB'_{n-1}}\right)\right\|_\alpha,\end{aligned} \tag{4.18}$$

where in the last equality, we define Θ by

$$\Theta_\sigma(\rho) \equiv \sigma^{1/2}\rho\sigma^{1/2}, \tag{4.19}$$

and the identity operation on the other systems is implied. We now use the key observation from Section 3.2 (and used in [BN05]) that if $\mathcal{N}$ is EB or if Alice uses separable inputs, throughout $\mathcal{P}_n$, Alice and Bob's systems are always separable. Furthermore, the M system is classical, so $\rho_{MA_nB'_{n-1}}$ is separable with respect to the $A_n : MB'_{n-1}$ cut. This implies that

$$\tau_{MB'_{n-1}}^{(1-\alpha)/(2\alpha)}\rho_{MA_nB'_{n-1}}\tau_{MB'_{n-1}}^{(1-\alpha)/(2\alpha)} \tag{4.20}$$

is a positive semidefinite separable operator:

$$\tau_{MB'_{n-1}}^{(1-\alpha)/(2\alpha)}\rho_{MA_nB'_{n-1}}\tau_{MB'_{n-1}}^{(1-\alpha)/(2\alpha)} = \tau_{MB'_{n-1}}^{(1-\alpha)/(2\alpha)}\left(\sum_j p(j)\rho^j_{A_n}\otimes\rho^j_{MB'_{n-1}}\right)\tau_{MB'_{n-1}}^{(1-\alpha)/(2\alpha)} \tag{4.21}$$

$$= \sum_j p(j)\rho^j_{A_n}\otimes\left(\tau_{MB'_{n-1}}^{(1-\alpha)/(2\alpha)}\rho^j_{MB'_{n-1}}\tau_{MB'_{n-1}}^{(1-\alpha)/(2\alpha)}\right). \tag{4.22}$$

Since conjugation by a positive semidefinite operator is clearly a completely positive map, we can apply Lemma 2 to conclude that

$$\left\|\left(\Theta_{\sigma_{B_n}^{\frac{1-\alpha}{\alpha}}}\circ\mathcal{N}_{A_n\to B_n}\right)\left(\tau_{MB'_{n-1}}^{(1-\alpha)/(2\alpha)}\rho_{MA_nB'_{n-1}}\tau_{MB'_{n-1}}^{(1-\alpha)/(2\alpha)}\right)\right\|_\alpha$$

$$\leqslant \nu_\alpha\left(\Theta_{\sigma_{B_n}^{\frac{1-\alpha}{\alpha}}}\circ\mathcal{N}_{A_n\to B_n}\right)\cdot\left\|\tau_{MB'_{n-1}}^{(1-\alpha)/(2\alpha)}\rho_{MB'_{n-1}}\tau_{MB'_{n-1}}^{(1-\alpha)/(2\alpha)}\right\|_\alpha. \tag{4.23}$$

We then have the following chain of inequalities:

$$\widetilde{D}_\alpha(\rho_{MB_nB'_{n-1}}\|\tau_{MB_nB'_{n-1}}) \leqslant \frac{\alpha}{\alpha-1}\log\nu_\alpha\left(\Theta_{\sigma_{B_n}^{\frac{1-\alpha}{\alpha}}}\circ\mathcal{N}\right) + \widetilde{D}_\alpha(\rho_{MB'_{n-1}}\|\tau_{MB'_{n-1}}) \tag{4.24}$$

$$\leqslant \frac{\alpha}{\alpha-1}\log\nu_\alpha\left(\Theta_{\sigma_{B_n}^{\frac{1-\alpha}{\alpha}}}\circ\mathcal{N}\right) + \widetilde{D}_\alpha(\rho_{MB_{n-1}B'_{n-2}}\|\tau_{MB_{n-1}B'_{n-2}}) \tag{4.25}$$

$$\leqslant n\frac{\alpha}{\alpha-1}\log\nu_\alpha\left(\Theta_{\sigma_{B_n}^{\frac{1-\alpha}{\alpha}}}\circ\mathcal{N}\right) + \widetilde{D}_\alpha(\rho_{MB'_0}\|\tau_{MB'_0}) \tag{4.26}$$

$$= n\frac{\alpha}{\alpha-1}\log\nu_\alpha\left(\Theta_{\sigma_{B_n}^{\frac{1-\alpha}{\alpha}}}\circ\mathcal{N}\right) \tag{4.27}$$

$$= n\frac{\alpha}{\alpha-1}\log\sup_{\rho_A\in\mathcal{S}(\mathcal{H}_A)}\left\|\left(\Theta_{\sigma_B^{\frac{1-\alpha}{\alpha}}}\circ\mathcal{N}_{A\to B}\right)(\rho_A)\right\|_\alpha. \tag{4.28}$$

The first inequality follows by combining (4.18) and (4.23). The second inequality follows from monotonicity of the sandwiched Rényi relative entropy under the partial trace channel. The third inequality follows by recognizing that $\widetilde{D}_\alpha(\rho_{MB_{n-1}B'_{n-2}}\|\tau_{MB_{n-1}B'_{n-2}})$ is the relative entropy at round $n-1$ of $\mathcal{P}_n$, which allows us to apply the above argument inductively. This is a crucial step of the

argument and proves a form of additivity similar to that of (1.5). The first equality is a consequence of the fact that $\rho_{MB_0'} = \tau_{MB_0'}$, since no channels have been applied at that point in the protocol. The last equality follows from a result from [Aud09] which allows us to take the supremum over quantum states instead of all trace class operators (furthermore from the fact that quantum states have trace equal to one). Hence, we have that

$$\frac{\alpha}{\alpha-1}\log p_{\text{succ}}(\mathcal{P}_n) + \log L \leqslant n\frac{\alpha}{\alpha-1}\log \sup_{\rho_A\in\mathcal{S}(\mathcal{H}_A)} \left\| \left(\Theta_{\sigma_B^{\frac{1-\alpha}{\alpha}}} \circ \mathcal{N}_{A\to B}\right)(\rho_A)\right\|_\alpha . \tag{4.29}$$

Since this is true for all states σ_B, we can conclude that

$$\begin{aligned}\frac{\alpha}{\alpha-1}\log p_{\text{succ}}(\mathcal{P}_n) + \log L &\leqslant n \inf_{\sigma_B\in\mathcal{S}(\mathcal{H}_B)} \sup_{\rho_A\in\mathcal{S}(\mathcal{H}_A)} \frac{\alpha}{\alpha-1}\log \left\| \left(\Theta_{\sigma_B^{\frac{1-\alpha}{\alpha}}} \circ \mathcal{N}_{A\to B}\right)(\rho_A)\right\|_\alpha && (4.30)\\ &= n\widetilde{K}_\alpha(\mathcal{N}) && (4.31)\\ &= n\widetilde{\chi}_\alpha(\mathcal{N}). && (4.32)\end{aligned}$$

The first equality comes from the definition of the α-information radius, and the second equality is due to Lemma 3.

Now, the protocol uses the channel n times and the rate R of $\mathcal{P}_n$ is defined to be the number of bits per channel use, so that $R \leqslant \frac{\log L}{n}$. This allows us to introduce R into the inequality:

$$\frac{1}{n}\log p_{\text{succ}}(\mathcal{P}_n) \leqslant -\frac{\alpha-1}{\alpha}(R - \widetilde{\chi}_\alpha(\mathcal{N})). \tag{4.33}$$

Since this is true for all $\alpha > 1$, we can take a supremum over $\alpha > 1$ and arrive at the bound stated in (4.10). □

The strong converse itself now follows from Theorem 5 and Lemma 4.

Corollary 6 (Strong Converse). *The probability of success of any sequence of protocols which use an entanglement-breaking channel with classical feedback at a rate greater than the classical capacity is bounded from above by a decaying exponential. The same is true for arbitrary channels with separable inputs.*

Proof. Recall that since $\widetilde{D}_\alpha$ is monotonically increasing in α, so is $\widetilde{\chi}_\alpha$. This along with Lemma 4 implies

$$\begin{aligned}\inf_{\alpha>1}\widetilde{\chi}_\alpha(\mathcal{N}) &= \lim_{\alpha\searrow 1}\widetilde{\chi}_\alpha(\mathcal{N}) && (4.34)\\ &= \chi(\mathcal{N}). && (4.35)\end{aligned}$$

Hence, if $R > \chi(\mathcal{N})$, by the continuity of $\widetilde{\chi}_\alpha(\mathcal{N})$ as $\alpha \searrow 1$, there exists a value of $\alpha > 1$ such that $\frac{\alpha-1}{\alpha}(R - \widetilde{\chi}_\alpha(\mathcal{N})) > 0$. Then the bound in (4.10) implies an exponential decay of the success probability. □

5 Conclusion

By studying the classical capacity of a quantum channel, quantum information theorists found that an implication of quantum mechanics for classical communication is the possibility of superior encoding schemes that use entanglement, a uniquely quantum phenomenon [Has09]. However, entanglement does not give an advantage for every channel. The results in this and previous papers show that for entanglement-breaking channels, no communication protocols, even with classical feedback, can take advantage of this extra resource. They cannot use it to increase their capacity, nor, as proved in this paper, to even lift the exponentially decaying ceiling on their success probabilities. We perhaps expect this since by definition EB channels destroy entanglement, and classical feedback channels cannot create entanglement. This is the guiding principle behind our proof, in particular the key fact that the transmitter and receiver states are separable throughout the protocol.

More generally, it is known that classical feedback can give a large boost to the classical capacity of channels which are not entanglement-breaking in at least two different ways: first, there exist channels for which the single-copy Holevo information is small but the single-copy Holevo information with feedback included can be quite large [BDSS06]. Similarly, there exist channels for which the classical capacity is small but becomes large when classical feedback is available [SS09]. In light of these prior results, this paper in some sense gives the strongest result that we could expect: going outside of the class of entanglement-breaking channels gives both nonadditivity and big gains from feedback. Thus, showing that the Holevo information is a strong-converse bound suggests that for feedback not to help, we require special channels, such as the entanglement-breaking ones.

Going in the opposite direction, it is known that EB channels form a proper superset of classical channels [KHH12]. We thus obtain the result of [PV10] as a direct corollary. Furthermore, since the original posting of our paper as arXiv:1506.02228, an open question that we posed has now been answered — the strong converse exponent from Theorem 5 is tight, due to the results in [MO15]. Hence, we obtain as a special case from this and our result the classical results of [Aug78, DK79, CK82] as well.

A possible direction for future research is to ask the same question for Hadamard channels, which are defined to be complements of EB channels. To define the complementary channel, we first explain the interpretation of channel as a model for open quantum dynamics. That is, a channel from system A to system B can be interpreted as the restriction of a unitary interaction with larger system BR, where R is referred to as the "environment." The complementary channel is the map from A to BR tracing out the B system instead and is itself a channel from A to R. Hence, Hadamard channels break any entanglement with the environment system that is traced out and is thus related to our guiding principle. The strong converse has already been proved for Hadamard channels [WWY14], but with the addition of classical feedback, even a weak converse has yet to be proved. Examples of Hadamard channels include generalized dephasing channels, cloning channels, and the Unruh channel [BHTW10].

Finally, we remark here that it should be possible to use the methods given here and in [CMW16] in order to characterize a particular adaptive hypothesis testing scenario. Suppose that the goal is to distinguish an entanglement-breaking channel from a replacement channel by means of adaptive, separability-preserving channels. Then the optimal strong converse exponent should be given in terms of a quantity similar to that in (4.10). However, we leave the details for future work.

Acknowledgements. We are grateful to Patrick Hayden, Milan Mosonyi, and Graeme Smith

for insightful discussions about the topic of this paper. We also thank an anonymous referee for many helpful suggestions for improving the paper. MMW acknowledges support from startup funds from the Department of Physics and Astronomy at LSU, the NSF under Award No. CCF-1350397, and the DARPA Quiness Program through US Army Research Office award W31P4Q-12-1-0019. DD acknowledges support from a Stanford Graduate Fellowship.

A Additivity

Here we justify the statement in (1.5). The backward implication is trivial. As for the forward implication, we first note that given n, for all $k \in \mathbb{Z}^+$,

$$\frac{1}{n}\chi(\mathcal{N}^{\otimes n}) \leqslant \frac{1}{kn}\chi(\mathcal{N}^{\otimes kn}), \tag{A.1}$$

because we can split the kn channels into k blocks of n channels and encode each block independently. In particular, $\frac{1}{n}\chi(\mathcal{N}^{\otimes n}) \geqslant \chi(\mathcal{N})$. To show the opposite inequality, we assume the contrapositive: $\frac{1}{n}\chi(\mathcal{N}^{\otimes n}) > \chi(\mathcal{N})$ for some n. However, this means

$$\frac{1}{kn}\chi(\mathcal{N}^{\otimes kn}) - \chi(\mathcal{N}) \geqslant \frac{1}{n}\chi(\mathcal{N}^{\otimes n}) - \chi(\mathcal{N}) > 0, \tag{A.2}$$

for all $k \in \mathbb{Z}^+$ and thus the limit in (1.4) does not converge to $\chi(\mathcal{N})$.

B Weak converse and finite bounds

We recall for motivation and the reader's convenience the argument for the weak converse from [BN05].

Theorem 7. *Let $\mathcal{P}_n$ be a protocol for classical feedback-assisted classical communication over an entanglement-breaking channel $\mathcal{N}$ such that it uses the channel n times, has communication rate R, and has average probability of decoding error ε. Then, it satisfies the following inequality:*

$$R \leqslant \chi(\mathcal{N}) + g(n, \varepsilon), \tag{B.1}$$

where $\chi(\mathcal{N})$ is the Holevo information of the channel and $g(n, \varepsilon)$ is a real valued function such that

$$\lim_{\varepsilon \searrow 0} \lim_{n \to \infty} g(n, \varepsilon) = 0. \tag{B.2}$$

The same is true for a protocol for communication over an arbitrary channel given that the encoder does not entangle inputs across different uses of the channel.

Proof. We use the notation in Section 3.2 and take a general approach that is, for instance, presented in Section 19.3.2 of [Wil11]. Let $\overline{\Phi}$ denote the following "shared randomness" state:

$$\overline{\Phi}_{M\hat{M}} = \frac{1}{|\mathcal{M}|}\sum_m |m\rangle\langle m|_M \otimes |m\rangle\langle m|_{\hat{M}}\,, \tag{B.3}$$

where $|\mathcal{M}|$ is the size of the message set $\mathcal{M}$. Now, suppose the information processing task is common randomness generation instead of classical communication. If we require the state ρ at the end of the protocol to be ε'-close to $\overline{\Phi}$

$$\|\overline{\Phi} - \rho\|_1 \leqslant \varepsilon', \tag{B.4}$$

the Fannes-Audenaert inequality for mutual information [Fan73, Aud07] gives

$$\begin{aligned} nR &= I(M;\hat{M})_{\overline{\Phi}} && \text{(B.5)}\\ &\leqslant I(M;\hat{M})_{\rho} + f(n,\varepsilon'), && \text{(B.6)} \end{aligned}$$

where $f(n,\varepsilon')$ is some continuous function of n and ε' with the property: $\lim_{\varepsilon' \searrow 0} \lim_{n\to\infty} \frac{1}{n} f(n,\varepsilon') = 0$. Continuing, we have

$$\begin{aligned} I(M;\hat{M})_{\rho} &\leqslant I(M;B_n B'_{n-1})_{\rho} && \text{(B.7)}\\ &= I(M;B'_{n-1})_{\rho} + I(M;B_n|B'_{n-1})_{\rho}, && \text{(B.8)} \end{aligned}$$

where we applied the data processing inequality and the chain rule for conditional quantum mutual information. We use the chain rule again to get

$$I(M;B_n|B'_{n-1})_{\rho} \leqslant I(MB'_{n-1};B_n)_{\rho}. \tag{B.9}$$

From Section 3.2, if $\mathcal{N}$ is entanglement-breaking or if Alice uses separable inputs, Alice and Bob's systems are always separable. Hence, the global state before the nth channel use can be written as

$$\begin{aligned} \rho_{MA'_nA_nB'_{n-1}} &= \sum_m p_M(m)|m\rangle\langle m|_M \otimes \rho^m_{A'_nA_nB'_{n-1}} && \text{(B.10)}\\ &= \sum_m p_M(m)\,|m\rangle\langle m|_M \otimes \sum_y p_{Y|M}(y|m)\rho^{m,y}_{A'_nA_n} \otimes \rho^{m,y}_{B'_{n-1}}. && \text{(B.11)} \end{aligned}$$

Hence, the state after the channel is given by

$$\rho_{MA'_nB_nB'_{n-1}} = \sum_m p_M(m)\,|m\rangle\langle m|_M \otimes \sum_y p_{Y|M}(y|m)\mathcal{N}_{A_n\to B_n}\left(\rho^{m,y}_{A'_nA_n}\right) \otimes \rho^{m,y}_{B'_{n-1}}. \tag{B.12}$$

We introduce an auxiliary system Y that labels the separable sum over the index y. That is, Y is chosen such that the global state is

$$\rho_{MYA'_nB_nB'_{n-1}} \equiv \sum_{m,y} p_M(m)\,p_{Y|M}(y|m)\,|m\rangle\langle m|_M \otimes |y\rangle\langle y|_Y \otimes \mathcal{N}_{A_n\to B_n}\left(\rho^{m,y}_{A'_nA_n}\right) \otimes \rho^{m,y}_{B'_{n-1}}. \tag{B.13}$$

We trace over A'_n to get

$$\rho_{MYB_nB'_{n-1}} = \sum_{m,y} p_M(m)\,p_{Y|M}(y|m)\,|m\rangle\langle m|_M \otimes |y\rangle\langle y|_Y \otimes \rho^{m,y}_{B_n} \otimes \rho^{m,y}_{B'_{n-1}}, \tag{B.14}$$

where

$$\rho^{m,y}_{B_n} \equiv \mathrm{Tr}_{A'_n}\left(\mathcal{N}_{A_n\to B_n}\left(\rho^{m,y}_{A'_nA_n}\right)\right) = \mathcal{N}_{A_n\to B_n}\left(\rho^{m,y}_{A_n}\right). \tag{B.15}$$

This allows us to argue

$$I\left(MB'_{n-1};B_n\right)_\rho \leqslant I\left(MYB'_{n-1};B_n\right)_\rho \tag{B.16}$$
$$= I\left(MY;B_n\right)_\rho + I\left(B'_{n-1};B_n|MY\right)_\rho \tag{B.17}$$
$$= I\left(MY;B_n\right)_\rho \tag{B.18}$$
$$\leqslant \chi(\mathcal{N}), \tag{B.19}$$

where the first inequality is from data processing, the first equality from the chain rule, and the second equality because the $B'_{n-1}B_n$ system is in a product state when conditioning on M and Y. Finally, given equation (B.15), the state ρ_{MYB_n} is a classical-quantum state of the form:

$$\rho_{MYB_n} = \mathrm{Tr}_{B'_{n-1}}\left(\rho_{MYB_nB'_{n-1}}\right) \tag{B.20}$$
$$= \sum_{m,y} p_M\left(m\right) p_{Y|M}(y|m)\left|m\right\rangle\!\left\langle m\right|_M \otimes \left|y\right\rangle\!\left\langle y\right|_Y \otimes \mathcal{N}_{A_n\to B_n}\left(\rho^{m,y}_{A_n}\right). \tag{B.21}$$

Hence, the definition of the Holevo information of the channel $\mathcal{N}$ gives us final inequality. Putting things together, we find that

$$I(M;\hat{M})_\rho \leqslant \chi(\mathcal{N}) + I\left(M;B'_{n-1}\right)_\rho \tag{B.22}$$
$$\leqslant \chi(\mathcal{N}) + I\left(M;B_{n-1}B'_{n-2}\right)_\rho, \tag{B.23}$$

where the last inequality follows from the data processing inequality. But now, we recognize the quantity $I\left(M;B_{n-1}B'_{n-2}\right)_\rho$ is of the same form as $I(M;B_nB'_{n-1})_\rho$ in (B.7), so that we can iterate through the same sequence of arguments to get

$$I(M;\hat{M})_\rho \leqslant 2\chi(\mathcal{N}) + I\left(M;B_{n-2}B'_{n-3}\right)_\rho. \tag{B.24}$$

Continuing all the way back to the first channel use, we find

$$I(M;\hat{M})_\rho \leqslant n\chi(\mathcal{N}) \tag{B.25}$$

since $I(M;B'_0) = 0$ (see (3.6)). Thus, we conclude

$$R \leqslant \chi(\mathcal{N}) + \frac{1}{n}f(n,\varepsilon'). \tag{B.26}$$

Now, the rate for classical communication with n channel uses and error ε is at most that of randomness generation with n uses and error ε', where ε' is some function of ε such that $\lim_{\varepsilon\searrow 0}\varepsilon' = 0$. Hence, $g(n,\varepsilon) \equiv \frac{1}{n}f(n,\varepsilon')$ is what we need. □

Corollary 8 (Weak Converse). *The classical capacity of entanglement-breaking channels with classical feedback is given by the Holevo information. The same is true for arbitrary channels with separable inputs.*

Proof. This is immediate. □

C Limit value of the α-Holevo information

The arguments here are essentially the same as those in [MH11, CMW16], along with an additional insight from [TWW14, Appendix A]. We first recall a minimax result due to [MH11, Corollary A2].

Lemma 9. *Suppose a function $f : X \times Y \to \bar{\mathbb{R}}$, where X is a compact topological space, Y is a subset of $\mathbb{R}$, and $\bar{\mathbb{R}}$ is the extended real numbers, satisfies*

1. *$\forall y \in Y$, $f(\cdot, y)$ is lower semicontinuous.*
2. *$\forall x \in X$, $f(x, \cdot)$ is monotonic.*

Then,

$$\inf_{x\in X} \sup_{y\in Y} f(x,y) = \sup_{y\in Y} \inf_{x\in X} f(x,y). \tag{C.1}$$

If the first condition was instead

1. *$\forall y \in Y$, $f(\cdot, y)$ is upper semicontinuous,*

then

$$\inf_{y\in Y} \sup_{x\in X} f(x,y) = \sup_{x\in X} \inf_{y\in Y} f(x,y). \tag{C.2}$$

We now prove Lemma 4.

Proof. For fixed $\rho_{XR'}$ and α, $\sigma_{R'} \mapsto \widetilde{D}_\alpha(\rho_{XR'}, \rho_X \otimes \sigma_{R'})$ is lower semicontinuous (see, e.g., [CMW16, Appendix A]). Furthermore, for fixed $\rho_{XR'}$ and $\sigma_{R'}$, $\alpha \mapsto \widetilde{D}_\alpha(\rho_{XR'}, \rho_X \otimes \sigma_{R'})$ is monotone non-decreasing. We can therefore invoke Lemma 9 with $X = \mathcal{S}(\mathcal{H}_{R'})$ and $Y = (0,1)$. We use this, the definition of the Holevo information, and properties of the quantum relative entropy to find the following one-sided limit of the α-Holevo information of a quantum channel $\mathcal{N} : \mathcal{T}(\mathcal{H}_R) \to \mathcal{T}(\mathcal{H}_{R'})$:

$$\begin{aligned}
\lim_{\alpha\nearrow 1} \widetilde{\chi}_\alpha(\mathcal{N}) &= \sup_{\alpha\in(0,1)} \widetilde{\chi}_\alpha(\mathcal{N}) && \text{(C.3)}\\
&= \sup_{\alpha\in(0,1)} \sup_{\{p_X(x),\rho_x\}} \widetilde{\chi}_\alpha(\{p_X(x), \mathcal{N}(\rho_x)\}) && \text{(C.4)}\\
&= \sup_{\{p_X(x),\rho_x\}} \sup_{\alpha\in(0,1)} \inf_{\sigma_{R'}\in\mathcal{S}(\mathcal{H}_{R'})} \widetilde{D}_\alpha(\rho_{XR'} \| \rho_X \otimes \sigma_{R'}) && \text{(C.5)}\\
&= \sup_{\{p_X(x),\rho_x\}} \inf_{\sigma_{R'}\in\mathcal{S}(\mathcal{H}_{R'})} \sup_{\alpha\in(0,1)} \widetilde{D}_\alpha(\rho_{XR'} \| \rho_X \otimes \sigma_{R'}) && \text{(C.6)}\\
&= \sup_{\{p_X(x),\rho_x\}} \inf_{\sigma_{R'}\in\mathcal{S}(\mathcal{H}_{R'})} D(\rho_{XR'} \| \rho_X \otimes \sigma_{R'}) && \text{(C.7)}\\
&= \sup_{\{p_X(x),\rho_x\}} I(X:R')_\rho && \text{(C.8)}\\
&= \chi(\mathcal{N}), && \text{(C.9)}
\end{aligned}$$

where

$$\rho_{XR'} \equiv \sum_x p_X(x) |x\rangle\langle x|_X \otimes [\mathcal{N}(\rho_x)]_{R'}\,. \tag{C.10}$$

The fourth equality follows from applying Lemma 9. The fifth follows from (3.15) and the fact that the sandwiched Rényi relative entropy is monotone non-decreasing in α.

To find the other limit, we note that for fixed ρ_R and $\sigma_{R'}$, $\alpha \mapsto \widetilde{D}_\alpha(\mathcal{N}(\rho_R)\|\sigma_{R'})$ is monotone non-decreasing. We then use an idea from [TWW14, Appendix A]. Note that for fixed α and positive definite $\sigma_{R'}$, $\rho_R \mapsto \widetilde{D}_\alpha(\mathcal{N}(\rho_R)\|\sigma_{R'})$ is continuous (and thus upper semicontinuous). Let $\varepsilon \in (0,1)$ and define $\sigma(\varepsilon) \equiv (1-\varepsilon)\sigma + \varepsilon\pi$, where π is the maximally mixed state. Thus $\sigma(\varepsilon)$ is positive definite for σ positive semidefinite. Consider that

$$\sigma(\varepsilon) \geqslant (1-\varepsilon)\sigma, \tag{C.11}$$

which implies that

$$D(\rho\|\sigma(\varepsilon)) \leqslant D(\rho\|(1-\varepsilon)\sigma) = D(\rho\|\sigma) - \log(1-\varepsilon). \tag{C.12}$$

(This is because of the well known fact that $\sigma \leqslant \sigma'$ implies that $D(\rho\|\sigma) \geqslant D(\rho\|\sigma')$.) Hence, using Lemma 9, Lemma 3, and the identification of the Holevo information as an information radius [OPW97, SW01], we can conclude the following chain of equalities:

$$\begin{aligned}
\lim_{\alpha\searrow 1} \widetilde{\chi}_\alpha(\mathcal{N}) &= \inf_{\alpha>1} \widetilde{\chi}_\alpha(\mathcal{N}) && \text{(C.13)}\\
&= \inf_{\alpha>1} \widetilde{K}_\alpha(\mathcal{N}) && \text{(C.14)}\\
&= \inf_{\alpha>1} \inf_{\sigma_{R'}\in\mathcal{S}(\mathcal{H}_{R'})} \sup_{\rho_R\in\mathcal{S}(\mathcal{H}_R)} \widetilde{D}_\alpha(\mathcal{N}(\rho_R)\|\sigma_{R'}) && \text{(C.15)}\\
&\leqslant \inf_{\alpha>1} \inf_{\sigma_{R'}\in\mathcal{S}(\mathcal{H}_{R'})} \sup_{\rho_R\in\mathcal{S}(\mathcal{H}_R)} \widetilde{D}_\alpha(\mathcal{N}(\rho_R)\|\sigma(\varepsilon)_{R'}) && \text{(C.16)}\\
&= \inf_{\sigma_{R'}\in\mathcal{S}(\mathcal{H}_{R'})} \inf_{\alpha>1} \sup_{\rho_R\in\mathcal{S}(\mathcal{H}_R)} \widetilde{D}_\alpha(\mathcal{N}(\rho_R)\|\sigma(\varepsilon)_{R'}) && \text{(C.17)}\\
&= \inf_{\sigma_{R'}\in\mathcal{S}(\mathcal{H}_{R'})} \sup_{\rho_R\in\mathcal{S}(\mathcal{H}_R)} \inf_{\alpha>1} \widetilde{D}_\alpha(\mathcal{N}(\rho_R)\|\sigma(\varepsilon)_{R'}) && \text{(C.18)}\\
&= \inf_{\sigma_{R'}\in\mathcal{S}(\mathcal{H}_{R'})} \sup_{\rho_R\in\mathcal{S}(\mathcal{H}_R)} D(\mathcal{N}(\rho_R)\|\sigma(\varepsilon)_{R'}) && \text{(C.19)}\\
&\leqslant \inf_{\sigma_{R'}\in\mathcal{S}(\mathcal{H}_{R'})} \sup_{\rho_R\in\mathcal{S}(\mathcal{H}_R)} D(\mathcal{N}(\rho_R)\|\sigma_{R'}) - \log(1-\varepsilon) && \text{(C.20)}\\
&= \chi(\mathcal{N}) - \log(1-\varepsilon). && \text{(C.21)}
\end{aligned}$$

The first inequality follows because there is an infimum with respect to $\sigma_{R'}$, such that replacing $\sigma_{R'}$ with $\sigma(\varepsilon)_{R'}$ can never decrease the quantity. The fifth equality follows from applying Lemma 9. The sixth equality follows from (3.15) and the fact that the sandwiched Rényi relative entropy is monotone non-decreasing in α. The last inequality follows from (C.12). The last equality uses $\chi(\mathcal{N}) = K(\mathcal{N})$.

Given that $\varepsilon > 0$ was arbitrary, we can conclude that $\lim_{\alpha\searrow 1} \widetilde{\chi}_\alpha(\mathcal{N}) \leqslant \chi(\mathcal{N})$. Combined with the fact that $\widetilde{\chi}_\alpha(\mathcal{N}) \geqslant \chi(\mathcal{N})$ for all $\alpha \geqslant 1$, we conclude that

$$\lim_{\alpha\searrow 1} \widetilde{\chi}_\alpha(\mathcal{N}) = \chi(\mathcal{N}). \tag{C.22}$$

This also implies that

$$\lim_{\alpha\searrow 1} \widetilde{K}_\alpha(\mathcal{N}) = K(\mathcal{N}). \tag{C.23}$$

We now show that $\lim_{\alpha \nearrow 1} \widetilde{K}_\alpha(\mathcal{N}) = K(\mathcal{N})$, which follows from a similar line of reasoning:

$$\lim_{\alpha \nearrow 1} \widetilde{K}_\alpha(\mathcal{N}) = \sup_{\alpha \in (0,1)} \widetilde{K}_\alpha(\mathcal{N}) \tag{C.24}$$

$$= \sup_{\alpha \in (0,1)} \inf_{\sigma_{R'} \in \mathcal{S}(\mathcal{H}_{R'})} \sup_{\rho_R \in \mathcal{S}(\mathcal{H}_R)} \widetilde{D}_\alpha(\mathcal{N}(\rho_R) \| \sigma_{R'}) \tag{C.25}$$

$$= \inf_{\sigma_{R'} \in \mathcal{S}(\mathcal{H}_{R'})} \sup_{\alpha \in (0,1)} \sup_{\rho_R \in \mathcal{S}(\mathcal{H}_R)} \widetilde{D}_\alpha(\mathcal{N}(\rho_R) \| \sigma_{R'}) \tag{C.26}$$

$$= \inf_{\sigma_{R'} \in \mathcal{S}(\mathcal{H}_{R'})} \sup_{\rho_R \in \mathcal{S}(\mathcal{H}_R)} \sup_{\alpha \in (0,1)} \widetilde{D}_\alpha(\mathcal{N}(\rho_R) \| \sigma_{R'}) \tag{C.27}$$

$$= \inf_{\sigma_{R'} \in \mathcal{S}(\mathcal{H}_{R'})} \sup_{\rho_R \in \mathcal{S}(\mathcal{H}_R)} D(\mathcal{N}(\rho_R) \| \sigma_{R'}) \tag{C.28}$$

$$= K(\mathcal{N}), \tag{C.29}$$

where the third equality follows from Lemma 9 and the fact that for fixed α and ρ_R, $\sigma_{R'} \mapsto \widetilde{D}_\alpha(\mathcal{N}(\rho_R) \| \sigma_{R'})$ is lower semicontinuous and that the pointwise supremum of lower semicontinuous functions is lower semicontinuous. □

References

[Ari73] Suguru Arimoto. On the converse to the coding theorem for discrete memoryless channels. *IEEE Transactions on Information Theory*, 19(3):357–359, May 1973.

[Aud07] Koenraad M. R. Audenaert. A sharp continuity estimate for the von Neumann entropy. *Journal of Physics A: Mathematical and Theoretical*, 40(28):8127, July 2007. arXiv:quant-ph/0610146.

[Aud09] Koenraad M. R. Audenaert. A note on the $p \to q$ norms of 2-positive maps. *Linear Algebra and its Applications*, 430(1):1436–1440, February 2009. arXiv:math-ph/0505085.

[Aug78] Udo Augustin. Noisy channels. Habilitation thesis, Universitat Erlangen-Nurnberg, West Germany, September 1978.

[BDSS06] Charles H. Bennett, Igor Devetak, Peter W. Shor, and John A. Smolin. Inequalities and separations among assisted capacities of quantum channels. *Physical Review Letters*, 96(15):150502, April 2006. arXiv:quant-ph/0406086.

[Bei13] Salman Beigi. Sandwiched Rényi divergence satisfies data processing inequality. *Journal of Mathematical Physics*, 54(12):122202, December 2013. arXiv:1306.5920.

[BGPWW15] Bhaskar Roy Bardhan, Raul Garcia-Patron, Mark M. Wilde, and Andreas Winter. Strong converse for the classical capacity of all phase-insensitive bosonic Gaussian channels. *IEEE Transactions on Information Theory*, 61(4):1842–1850, April 2015. arXiv:1401.4161.

[BHTW10] Kamil Brádler, Patrick Hayden, Dave Touchette, and Mark M. Wilde. Trade-off capacities of the quantum Hadamard channels. *Physical Review A*, 81(6):062312, June 2010. arXiv:1001.1732.

[BN05] Garry Bowen and Rajagopal Nagarajan. On feedback and the classical capacity of a noisy quantum channel. *IEEE Transactions on Information Theory*, 51(1):320–324, 2005. arXiv:quant-ph/0305176.

[CK82] Imre Csiszar and Janos Korner. Feedback does not affect the reliability function of a DMC at rates above capacity. *IEEE Transactions on Information Theory*, 28(1):92–93, January 1982.

[CMW16] Tom Cooney, Milan Mosonyi, and Mark M. Wilde. Strong converse exponents for a quantum channel discrimination problem and quantum-feedback-assisted communication. *Communications in Mathematical Physics*, 344(3):797–829, June 2016. arXiv:1408.3373.

[CT91] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, August 1991.

[DK79] Gunther Dueck and Janos Korner. Reliability function of a discrete memoryless channel at rates above capacity. *IEEE Transactions on Information Theory*, 25(1):82–85, January 1979.

[Fan73] Mark Fannes. A continuity property of the entropy density for spin lattice systems. *Communications in Mathematical Physics*, 31(4):291–294, December 1973.

[FL13] Rupert L. Frank and Elliott H. Lieb. Monotonicity of a relative Rényi entropy. *Journal of Mathematical Physics*, 54(12):122201, December 2013. arXiv:1306.5358.

[GK12] Abbas El Gamal and Young-Han Kim. *Network information theory*. Cambridge University Press, January 2012. arXiv:1001.3404.

[Has09] Matthew B. Hastings. Superadditivity of communication capacity using entangled inputs. *Nature Physics*, 5:255–257, April 2009. arXiv:0809.3972.

[Hay07] Masahito Hayashi. Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding. *Physical Review A*, 76(6):062301, December 2007. arXiv:quant-ph/0611013.

[Hol98] Alexander S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Transactions on Information Theory*, 44(1):269–273, January 1998. arXiv:quant-ph/9611023.

[Hol06] Alexander S. Holevo. Multiplicativity of p-norms of completely positive maps and the additivity problem in quantum information theory. *Russian Mathematical Surveys*, 61(2):301–339, 2006.

[HSR03] Michał Horodecki, Peter W. Shor, and Mary Beth Ruskai. Entanglement breaking channels. *Reviews in Mathematical Physics*, 15(6):629–641, August 2003. arXiv:quant-ph/0302031.

[KHH12] J. K. Korbicz, P. Horodecki, and R. Horodecki. Quantum-correlation breaking channels, broadcasting scenarios, and finite Markov chains. *Physical Review A*, 86(4):042319, October 2012. arXiv:1208.2162.

[Kin03] Christopher King. Maximal p-norms of entanglement breaking channels. *Quantum Information and Computation*, 3(2):186–190, 2003. arXiv:quant-ph/0212057.

[KW09] Robert Koenig and Stephanie Wehner. A strong converse for classical channel coding using entangled inputs. *Physical Review Letters*, 103(7):070504, August 2009. arXiv:0903.2838.

[MH11] Milán Mosonyi and Fumio Hiai. On the quantum Rényi relative entropies and related capacity formulas. *IEEE Transactions on Information Theory*, 57(4):2474–2487, April 2011. arXiv:0912.1286.

[MLDS+13] Martin Müller-Lennert, Frédéric Dupuis, Oleg Szehr, Serge Fehr, and Marco Tomamichel. On quantum Rényi entropies: A new generalization and some properties. *Journal of Mathematical Physics*, 54(12):122203, December 2013. arXiv:1306.3142.

[MO15] Milán Mosonyi and Tomohiro Ogawa. Strong converse exponent for classical-quantum channel coding. July 2015. arXiv:1409.3562v5.

[Nag01] Hiroshi Nagaoka. Strong converse theorems in quantum information theory. *Proceedings of ERATO Workshop on Quantum Information Science*, page 33, 2001. Also appeared in Asymptotic Theory of Quantum Statistical Inference, ed. M. Hayashi, World Scientific, 2005.

[ON99] Tomohiro Ogawa and Hiroshi Nagaoka. Strong converse to the quantum channel coding theorem. *IEEE Transactions on Information Theory*, 45(7):2486–2489, November 1999. arXiv:quant-ph/9808063.

[OPW97] Masanori Ohya, Denes Petz, and Noburu Watanabe. On capacities of quantum channels. *Probability and Mathematical Statistics-Wroclaw University*, 17:179–196, 1997.

[PV10] Yury Polyanskiy and Sergio Verdú. Arimoto channel coding converse and Rényi divergence. *Proceedings of the 48th Annual Allerton Conference on Communication, Control, and Computation*, pages 1327–1333, September 2010.

[Sho02] Peter W. Shor. Additivity of the classical capacity of entanglement-breaking quantum channels. *Journal of Mathematical Physics*, 43(9):4334–4340, October 2002. arXiv:quant-ph/0201149.

[SS09] Graeme Smith and John A. Smolin. Extensive nonadditivity of privacy. *Physical Review Letters*, 103(12):120503, September 2009. arXiv:0904.4050.

[SW97] Benjamin Schumacher and Michael Westmoreland. Sending classical information via noisy quantum channels. *Physical Review A*, 56(1):131–138, July 1997.

[SW01] Benjamin Schumacher and Michael D Westmoreland. Optimal signal ensembles. *Physical Review A*, 63(2):022308, January 2001. arXiv:quant-ph/9912122.

[SW02] Benjamin Schumacher and Michael D Westmoreland. Relative entropy in quantum information theory. *Contemporary Mathematics*, 305:265–290, 2002. arXiv:quant-ph/0004045.

[TWW14] Marco Tomamichel, Mark M. Wilde, and Andreas Winter. Strong converse rates for quantum communication. June 2014. arXiv:1406.2946.

[Ume62] Hisaharu Umegaki. Conditional expectation in an operator algebra. *Kodai Mathematical Seminar Reports*, 14(2):59–85, 1962.

[Wil11] Mark M. Wilde. From classical to quantum Shannon theory. June 2011. arXiv:1106.1445.

[Win99] Andreas Winter. Coding theorem and strong converse for quantum channels. *IEEE Transactions on Information Theory*, 45(7):2481–2485, 1999. arXiv:1409.2536.

[Wol78] Jacob Wolfowitz. *Coding Theorems of Information Theory*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 3rd edition, 1978.

[WWY14] Mark M. Wilde, Andreas Winter, and Dong Yang. Strong converse for the classical capacity of entanglement-breaking and Hadamard channels via a sandwiched Rényi relative entropy. *Communications in Mathematical Physics*, 331(2):593–622, October 2014. arXiv:1306.1586.